Community Detection in the Twitter Network of the U.S. Congress using the Louvain Algorithm

1st Abror Muhammad Hazim
School Of Computing
Telkom University
Bandung, Indonesia
abrormhazim@student.telkomuniversity.ac.id

2nd Armand Aryasatya Prakarsa
School Of Computing
Telkom University
Bandung, Indonesia
armandarsasp@student.telkomuniversity.ac.id

3rd Ikhwan Wahyudin
School Of Computing
Telkom University
Bandung, Indonesia
ikhwanwahyudin@student.telkomuniversity.ac.id

4th Warih Maharani
School Of Computing
Telkom University
Bandung, Indonesia
wmaharani@telkomuniversity.ac.id

***Abstract*—This research focuses on analyzing the interaction network on Twitter among members of the US Congress using the dataset "Twitter Interaction Network for the US Congress." The dataset includes 475 nodes representing members of Congress, with 13,289 directed interactions encompassing retweets, replies, and mentions. The primary objective of this study is to identify community structures within the network by employing the Louvain method for community detection. This approach allows for the grouping of members based on their interaction patterns and highlights cohesive subgroups within the network. The results reveal four distinct communities within the U.S. Congress network, illustrating the communication dynamics among members. These findings offer insights into their interaction patterns and potential alliances, with a modularity score of 0.4375 indicating a moderately strong community structure.**

Keywords— ***Community Detection, Louvain Method, Twitter, US Congress, Social Network Analysis.***

I. INTRODUCTION

Twitter has become a transformative tool in political engagement, shaping how U.S. Congress members communicate and build connections with the public. During the 2016 U.S. presidential election, Twitter proved instrumental for candidates to engage with voters directly, influence public opinion, and amplify their messages through rapid and accessible digital channels. This shift illustrates Twitter's powerful role in influencing public discourse and opinion formation [1].

Beyond presidential campaigns, Twitter enables Congress members to communicate policy positions and connect with their constituents effectively. Research indicates that Congressional outreach on Twitter not only enhances political engagement but also reaches underrepresented communities, such as Latino populations, by tailoring messages to specific groups [2]. The platform's interactive features, like retweets, replies, and mentions, create a dynamic network were information spreads quickly. This network structure, however, also fosters echo chambers, where ideologically similar users interact within their own groups, reinforcing preexisting biases and potentially intensifying polarization [3].

Political interactions on Twitter have significant implications for democratic engagement, potentially increasing civic participation by making citizens feel more connected and informed. This heightened sense of immediacy and accessibility fosters direct communication between representatives and their constituents, encouraging individuals to vote and participate in political processes [4].

This study aims to analyze the influence of U.S. Congress members within Twitter's information dissemination network, using the "Twitter Interaction Network for the US Congress" dataset from the Stanford Large Network Dataset Collection (SNAP). With 475 nodes representing Congress members and 13,289 directed interactions, this dataset allows for a detailed exploration of political communication patterns. By applying network analysis methods—focusing on community detection using modularity score and applying Louvain methods, this approach allows for the grouping of members based on their interaction patterns and highlights cohesive subgroups within the network. and examine how information flow affects political discourse on social media.

II. LITERATURE REVIEW

Community detection plays a pivotal role in analyzing social networks, with the Louvain algorithm being one of the most widely utilized methods due to its modularity optimization approach.

Sliwa et al. (2024) demonstrated the effectiveness of the Louvain algorithm in analyzing Twitter communication during the early stages of the Ukraine war in 2022. The algorithm identifies communities by maximizing modularity, focusing on dense intra-community connections while minimizing inter-community links. However, its tendency to create overly large communities that lack cohesion was highlighted as a limitation [15].

Similarly, Tekin and Bostanoğlu (2024) emphasized the Louvain algorithm's sensitivity to network modifications in the context of community detection attacks. By strategically adding or removing edges, these attacks can significantly alter the detected community structure, underscoring the algorithm's susceptibility to manipulation and the need for privacy-preserving methods [16].

III. METHODOLOGY

A. Sosial Network Analysis

The SNA is the most widely used method for finding experts on community-driven knowledge-sharing sites [12]. SNA is one type of method to analyze the attributes of actor interaction, with its application must meet the following conditions: First, the phenomenon must be explained through the attributes of the network structure, second, the interaction between entities can form a network. If it does not meet both conditions, then SNA is difficult to implement.[13]

B. Community Detection

Community detection is the process of identifying groups within a network where nodes are densely connected internally but sparsely connected externally. The Louvain algorithm is a modularity-based method widely recognized for its efficiency and scalability in large networks. It works iteratively in two phases: first, by locally optimizing modularity where each node is assigned to the community that maximizes modularity gain; and second, by aggregating the identified communities into single nodes to form a smaller graph. These steps are repeated until no further modularity improvement is possible, resulting in a final partition of the network that reflects its inherent community structure [5].

C. Louvain Method

The Louvain algorithm is an efficient hierarchical clustering method used for community detection in large networks. It maximizes modularity, a measure of the quality of a network's division into communities. The algorithm operates in two main phases: a local optimization phase and an aggregation phase. In the local optimization phase, each node is initially treated as its own community, and nodes are iteratively moved to neighboring communities if such a move increases modularity. This process continues until no further improvement is possible. In the aggregation phase, communities identified in the first phase are collapsed into single nodes, and the process is repeated on the newly formed network. These steps iterate until modularity reaches a maximum, resulting in a final partition of the network that reflects its underlying community structure [6].

D. Louvain Method

Algorithm 1. *Louvain Method*

```
from community import community_louvain

# Apply Louvain algorithm for community detection
partition = community_louvain.best_partition(nx.Graph(G))

# Assign community labels to nodes
nx.set_node_attributes(G, partition, 'community')

# Analyze community structure
community_characteristics = {}
for node, community_id in partition.items():
    if community_id not in community_characteristics:
        community_characteristics[community_id] = {
            'members': [],
            'size': 0
        }

    username = nodes_df.loc[node, 'username']
    community_characteristics[community_id]['members'].append(username)
    community_characteristics[community_id]['size'] += 1
```

This research uses Louvain's algorithm to detect communities in network graphs. This algorithm is a hierarchical clustering algorithm based on graph theory. The principle of Louvain's algorithm is to make the modularity of the maximum value community partition results through continuous iteration of moving nodes and the results will obtain the optimal community partition.

Louvain algorithm steps:

1. Initiate the community and assign each node to a separate community.
2. Search for all connected communities and calculate their modularity, if when the node is moved and the modularity increases then the node is moved to that community if not, then the node remains in the previous community.
3. Repeat on all nodes and apply Step 2 until no nodes need to be moved and we get a community partition layer.
4. Merge each community into a new node, and go back to step 1 until all nodes are finally merged into one community.

Here is the formula for the Louvain algorithm:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta\left(C_i C_j\right) \quad \text{...(1)}$$

k_i = is the degree of vertex i
c_i = is the sum of the weights of all edges in community i
A_{ij} = is the weight of the edge connected between nodes i and j
$\delta(C_i C_j)$ = represents whether nodes i and j are in the same community.
$\delta(C_i C_j) = 1$ = 1 if Yes is in a community, and 0 otherwise.

E. Twitter

X or Twitter is increasingly used in everyday life by people regardless of one's geographic location. Approximately 554.7 million people worldwide actively use it, posting approximately 58 million "tweets" every day, and 135,000 new users join the X or twitter app.[14]

PHASES OF RESEARCH

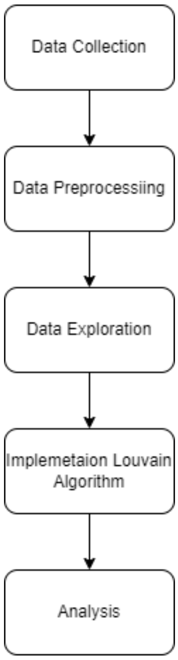

Fig 1. Phases of Research Process

This research phase will analyze the community detection using the Louvain algorithm on the "Twitter Interaction Network for the US Congress" dataset from the Standford Network Analysis Project (SNAP).

F. Data collection

This study uses the “Twitter Interaction Network for the US Congress” dataset, downloaded from the Standford Network Analysis Project (SNAP). This dataset consists of 475 nodes (congressmen) and 13,289 interactions (edges) covering three main types of interactions, namely retweets, mentions, and quotes, taken from the official Twitter accounts of senators and representatives.

TABLE 1. Dataset Statistics

Dataset Statistics	
Directed	Yes
Node Features	No
Edge Features	Yes
Nodes	475
Edges	13.289
Format	Json

G. Graph construction formation

Raw data in JSON format is processed to form a directed graph structure, where nodes represent congress members and directed edges represent interactions. Duplicate edges are removed, and isolated nodes are retained to preserve the dataset's integrity.

H. Graph visualization

The graph is visualized using the NetworkX library in Python to understand the network's structure. The visualization highlights nodes, edges, and clusters, providing insights into the relationships between members.

I. Implementation of Louvain Algorithm

Louvain algorithm is applied to the graph to detect communities. This algorithm iteratively optimizes modularity, a measure of the strength of division of a network into communities. The algorithm assigns a community label to each node, identifying clusters with high intra-cluster connectivity.

J. Results Analysis

After all methods are applied, the next step is to evaluate and analyze the results of the Louvain algorithm.

IV. Result and Conclusion

A. Dataset

TABLE 2. Preview Dataset

Source	Target	Weight	Source Username	Target Username
4	0	0.0036496350364963502	SenBlumenthal	SenatorBaldwin
9	0	0.0036101083032490976	SenSherrodBrown	SenatorBaldwin
11	0	0.004073319755600814	SenCapito	SenatorBaldwin
13	0	0.0031446540880503146	SenatorCarper	SenatorBaldwin
18	0	0.002347417840375587	SenCortezMasto	SenatorBaldwin

The table presented displays a simplified version of the dataset. The original dataset contains more detailed information, which is structured as follows:

- inList: A list of lists where inList[i] contains all the nodes that send connections to node i.
- inWeight: A list of lists that holds the connection weights (transmission probabilities) corresponding to the connections in inList.
- outList: A list of lists where outList[i] contains all the nodes that receive connections from node i.
- outWeight: A list of lists that contains the connection weights (transmission probabilities) corresponding to the connections in outList.
- usernameList[i] provides the Twitter username associated with node i.

B. Data Preprocessing

During data preprocessing phase, we examined the dataset for missing values and identified the data types of each column. The following summary provides an overview of the dataset structure after preprocessing:

TABLE 3. Data Information

#	Column	Non-Null Count	Type
0	username	475 non-null	object
1	In_connection	475 non-null	object
2	Out_connection	475 non-null	object

As shown, all columns contain 475 non-null entries, indicating that there are no missing values in the dataset. The data type for each column is identified as object, which typically represents string or categorical data.

C. Data Exploratory

TABLE 4. Degree Summary

Degree Type	Count	Mean	Std Dev	Min	25%	50%	75%	Max
in_degree	475	27.98	0,94375	0	13	22	37	127
out_degree	475	27.98	18.35	1	17	24	35	210
total_degree	475	55.95	34.83	2	33	48	69	284

The degree summary statistics provide insights into the distribution of node connections in the dataset. The mean in-degree and out-degree are both around 27.98, with a total-degree mean of 55.95. The high standard deviations (21.99 for in-degree, 18.35 for out-degree) suggest a wide variation in the number of connections, with a few nodes having significantly higher degrees. The maximum total-degree is 284, indicating some nodes with a much higher number of connections than others.

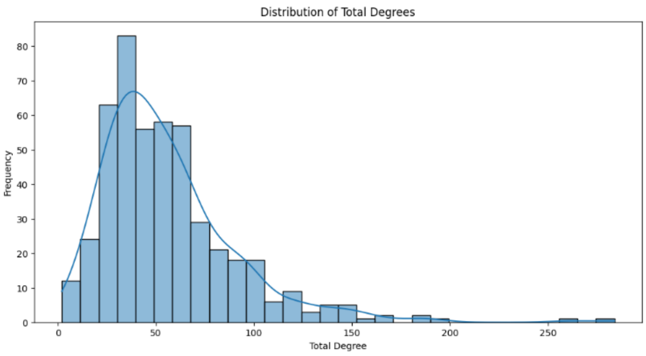

Fig 2. Degree Distribution

The histogram of total-degree shows a positively skewed distribution, with most nodes having a degree between 30-50. A small number of nodes have much higher degrees, peaking above 100, indicating a few highly connected members with broad networks. This long tail suggests that some members of Congress play a more central role in the network, likely influencing discussions and information flow. This degree distribution is typical of social networks, where most individuals have moderate connections, but a few "hubs" dominate.

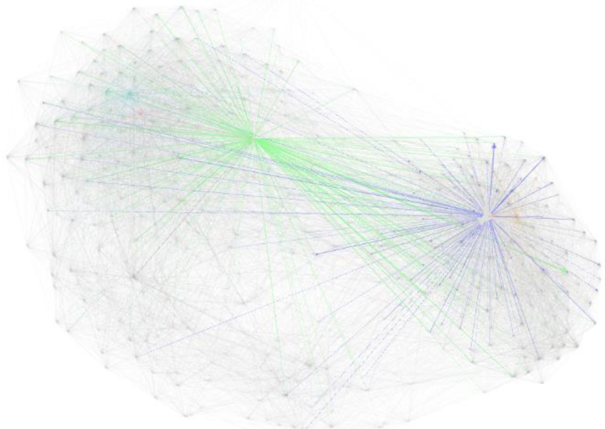

Fig 3. Network visualization

The implementation of the Louvain method algorithm for community detection in a graph is carried out using the Python library community. The process begins with importing the *community_louvain* module and loading the graph to be analyzed using NetworkX. The Louvain algorithm is applied through the *community_louvain.best_partition()* function, which returns a dictionary where each key is the node ID, and the value is the assigned community ID. These community attributes are then set to the nodes in the graph using *nx.set_node_attributes()*, enabling the identification of communities for each node. Further analysis of the community structure is performed by creating a *community_characteristics* dictionary, which stores the members and size of each community. This process is done by iterating over the partition results and adding user information from the nodes_df DataFrame to the relevant community. Each community is identified by its ID, and the members and size of the community are updated dynamically. The results of this implementation can be used to study community composition and evaluate the existing network structure, thus providing deeper insights into the relationships between nodes in the graph.

TABLE 5. Preview Community Characteristics

Community	Total Members	Preview Members
1	175	Robert_Aderholt, RepRickAllen, RepArmstrongND
2	93	SenatorBaldwin, SenJohnBarrasso, SenatorBennet
3	186	SenatorCantwell, SenDuckworth, SenMarkey
4	21	SenatorCardin, ChrisVanHollen, MarkWarner

The analysis revealed four distinct communities, each varying in size and potentially reflecting underlying groupings within the network. Specifically, Community 1 comprised 175 members, Community 2 included 93 members, Community 3 had 186 members, and Community 4 was the smallest with 21 members. The modularity score of the resulting partition was calculated to be 0.4375, which indicates a moderately strong community structure, suggesting that the graph is effectively segmented into well-defined subgroups.

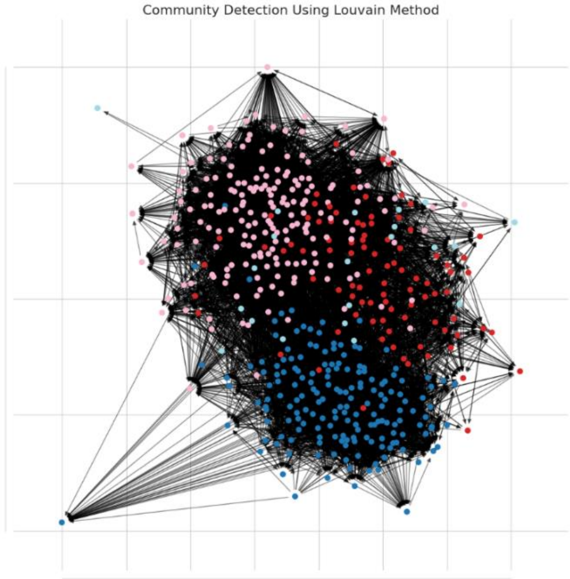

Fig 4. Community Detection Visualization

V. CONCLUSION

Based on the results of experiments the application of the Louvain algorithm for community detection on the constructed graph has provided valuable insights into the network's structure. The algorithm successfully partitioned the graph into four distinct communities, with varying sizes, and achieved a modularity score of 0.4375, indicating a moderately strong community structure. This finding suggests that the network, which represents the U.S. Congress, is composed of subgroups that could be influenced by factors such as political party affiliations, policy interests, or levels of member interaction. The identification of these communities is significant as it enhances our understanding of the complex relationships and collaborative dynamics within the legislative body. Future research could explore the characteristics and interactions within these subgroups in greater depth, examining how these community structures influence decision-making, policy formulation, and overall legislative processes. This study lays the groundwork for further analysis into the political and social structures that drive collaboration and competition in the context of U.S. governance.

REFERENCES

[1] Buccoliero, L., Bellio, E., Crestini, G., & Arkoudas, A. (2018). Twitter and politics: Evidence from the US presidential elections 2016. Journal of Marketing Communications, 1–27.

[2] Gervais, B. T., & Wilson, W. C. (2017). New media for the new electorate? Congressional outreach to Latinos on Twitter. Politics, Groups, and Identities, 1–19. doi:10.1080/21565503.2017.1358186

[3] Guo, L., Rohde, J. A., & Wu, H. D. (2018). Who is responsible for Twitter's echo chamber problem? Evidence from 2016 U.S.

election networks. Information, Communication & Society, 1–18. doi:10.1080/1369118x.2018.1499793

[4] T. Rotesi, "The Impact of Twitter on Political Participation," University of Lausanne, Department of Economics, 2018.

[5] M. Seifikar, S. Farzi, and M. Barati, “C-Blondel: An Efficient Louvain-Based Dynamic Community Detection Algorithm,” IEEE Transactions on Computational Social Systems, Vol. 7, no.2, pp. 308-318

[6] J. Zhang, J. Fei, X. song, and J. Feng, “An Improved Louvain Algorithm for Community Detection Mathematical Problems in Engineering, vol. 2021, Article ID 1485592, pp. 1-4, Nov. 2021

[7] Ledesma González, O., Merinero-Rodríguez, R., & Pulido-Fernández, J. I. (2021). Tourist destination development and social network analysis: What does degree centrality contribute? International Journal of Tourism Research, 23(4), 652–666. doi:10.1002/jtr.2432

[8] Rodrigues, F.A. (2019). Network Centrality: An Introduction. In: Macau, E. (eds) A Mathematical Modeling Approach from Nonlinear Dynamics to Complex Systems . Nonlinear Systems and Complexity, vol 22. Springer, Cham. https://doi.org/10.1007/978-3-319-78512-7_10

[9] D.P Kingma and M. Welling, “Auto-encoding variational Bayes,” 2013, arXiv:1312.6114. [Online]. Available: https://arxiv.org/abs/1312.6114

[10] S. Liu, “Wi-Fi Energy Detection Testbed (12MTC),” 2023, gitHub repository. [Online]. Available: https://github.com/liustone99/Wi-Fi-Energy-Detection-Testbed-12MTC

[11] “Treatment episode data set: discharges (TEDS-D): concatenated, 2006 to 2009.” U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies, August, 2013, DOI:10.3886/ICPSR30122.v2

[12] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. (references)

[13] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[14] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[15] K. Sliwa, E. Kušen, and M. Strembeck, "A Case Study Comparing Twitter Communities Detected by the Louvain and Leiden Algorithms During the 2022 War in Ukraine," in *Companion Proceedings of the ACM Web Conference 2024 (WWW ’24 Companion)*, Singapore, Singapore, May 2024, pp. 1376–1380. doi: 10.1145/3589335.3651892.

[16] L. Tekin and B. E. Bostanoğlu, "A Qualitative Survey on Community Detection Attack Algorithms," *Symmetry*, vol. 16, no. 10, p. 1272, Sep. 2024. doi: 10.3390/sym16101272.